U-STATISTICS AND RANDOM SUBGRAPH COUNTS: MULTIVARIATE NORMAL APPROXIMATION VIA EXCHANGEABLE PAIRS AND EMBEDDING

GESINE REINERT AND ADRIAN RÖLLIN

University of Oxford and National University of Singapore

Abstract. In a recent paper [8] a new approach—called the "embedding method"—was introduced, which allows to make use of exchangeable pairs for normal and multivariate normal approximation with Stein's method in cases where the corresponding couplings do not satisfy a certain linearity condition. The key idea is to embed the problem into a higher dimensional space in such a way that the linearity condition is then satisfied. Here we apply the embedding to U-statistics as well as to subgraph counts in random graphs.

1. Introduction

Stein's method, first introduced in the 70s [11], has proven a powerful tool for assessing distributional distances, such as to the normal distribution, in the presence of dependence. When considering sums W of random variables, the dependence between these random variables needs to be weak in order for the distance to a normal distribution to be small. For quantifying weak dependence, Stein [12] introduced the method of exchangeable pairs: construct a sum W' such that (W, W') form an exchangeable pair, and such that $\mathbb{E}^W(W' - W)$ is (at least approximately) linear in W. This linearity condition arises naturally when thinking of correlated bivariate normals. The generalisation of this approach to a multivariate setting remained untackled until recently Chatterjee and Meckes [2] solved the problem in the case of exchangeable vectors (W, W') such that $\mathbb{E}^W(W' - W) = -\lambda W + R$ for a scalar λ and a remainder vector R such that $\mathbb{E}|R|$ is small. This is a rather special case; the authors [8] tackled the general setting that

$$\mathbb{E}^W(W' - W) = -\Lambda W + R \tag{1.1}$$

for a matrix Λ and a vector R with small $\mathbb{E}|R|$ is treated. In in a followup paper by Meckes [5] the results by Chatterjee and Meckes [2] and by the authors [8] are combined using slightly different smoothness conditions on test functions as compared to [8]; non-smooth test functions are not treated by Meckes [5], but the bounds obtained there improve on those from [8] for the example of d-runs with respect to smooth test functions.

A surprising finding in [8] was that it is often possible to embed a random vector W into a random vector $\hat{W}$ of larger, but still finite, dimension, such that (1.1) holds with $R = 0$; yet this embedding does not correspond to Hoeffding projections. Here we explore the embedding method further, by illustrating its use on two important examples. The first example is complete non-degenerate U-statistics, and the second example considers the

joint count of edges and triangles in Bernoulli random graphs. In both examples the limiting covariance matrix is not of full rank; yet the bounds on the normal approximation are of the expected order.

The paper is organised as follows. In Section 2 we review the theoretical results in [8], giving bounds on the distance to normal under the linearity condition (1.1), both for smooth test functions and for non-smooth test functions. In Section 3 we discuss the embedding method, and point out a link to Rademacher integrals and chaos decompositions. Section 4 illustrates the embedding method for complete non-degenerate U-statistics; the embedding vector contains lower-order U-statistics which are obtained via fixing components. Section 5 gives a normal approximation for the joint counts of the number of edges and the number of triangles in a Bernoulli random graph; to our knowledge these are the first explicit bounds for this multivariate problem. The embedding method suggests to count the number of 2-stars as well, which makes the results not only more informative but also, surprisingly, easier to derive.

2. Theoretical bounds for a multivariate normal approximation

2.1. **Notation.** Denote by $W = (W_1, W_2, \dots, W_d)^t$ random vectors in $\mathbb{R}^d$, where W_i are $\mathbb{R}$-values random variables for $i = 1, \dots, d$. We denote by Σ symmetric, non-negative definite matrices, and hence by $\Sigma^{1/2}$ the unique symmetric square root of Σ. Denote by Id the identity matrix, where we omit the dimension d. Throughout this article, Z denotes a random variable having standard d-dimensional multivariate normal distribution. We abbreviate the transpose of the inverse of a matrix Λ as $\Lambda^{-t} := (\Lambda^{-1})^t$.

For derivatives of smooth functions $h : \mathbb{R}^d \to \mathbb{R}$, we use the notation ∇ for the gradient operator. Denote by $\|\cdot\|$ the supremum norm for both functions and matrices. If the corresponding derivatives exist for some function $h : \mathbb{R}^d \to \mathbb{R}$, we abbreviate $|h|_1 := \sup_i \bigl\| \frac{\partial}{\partial x_i} h \bigr\|$, $|h|_2 := \sup_{i,j} \bigl\| \frac{\partial^2}{\partial x_i \partial x_j} h \bigr\|$, and so on.

We start by considering smooth test functions.

Theorem 2.1 (c.f. Theoem 2.1 [8])**.** *Assume that (W, W') is an exchangeable pair of $\mathbb{R}^d$-valued random variables such that*

$$\mathbb{E}W = 0, \qquad \mathbb{E}WW^t = \Sigma, \tag{2.1}$$

with $\Sigma \in \mathbb{R}^{d\times d}$ symmetric and positive definite. Suppose further that (1.1) is satisfied for an invertible matrix Λ and a $\sigma(W)$-measurable random variable R. Then, if Z has d-dimensional standard normal distribution, we have for every three times differentiable function h,

$$\bigl|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2}Z)\bigr| \leqslant \frac{|h|_2}{4}A + \frac{|h|_3}{12}B + \Bigl(|h|_1 + \tfrac{1}{2}d\|\Sigma\|^{1/2}|h|_2\Bigr)C, \tag{2.2}$$

where, with $\lambda^{(i)} = \sum_{m=1}^d |(\Lambda^{-1})_{m,i}|$,

$$A = \sum_{i,j=1}^d \lambda^{(i)} \sqrt{\operatorname{Var} \mathbb{E}^W (W_i' - W_i)(W_j' - W_j)},$$

$$B = \sum_{i,j,k=1}^d \lambda^{(i)} \mathbb{E}\big|(W_i' - W_i)(W_j' - W_j)(W_k' - W_k)\big|,$$

$$C = \sum_i \lambda^{(i)} \sqrt{\mathbb{E} R_i^2}.$$

The proof of Theorem 2.1 is based on the Stein characterization of the normal distribution that $Y \in \mathbb{R}^d$ is a multivariate normal $\mathrm{MVN}(0, \Sigma)$ if and only if

$$\mathbb{E} Y^t \nabla f(Y) = \mathbb{E} \nabla^t \Sigma \nabla f(Y), \quad \text{for all smooth } f : \mathbb{R}^d \to \mathbb{R}. \tag{2.3}$$

In the paper by Meckes [5] a different norm for functions and for operators is used, to obtain a similar result, and the difference in the bounds depending on the chosen norm is illustrated for the example of runs on the line.

Theorem 2.1 can be extended to allow for covariance matrices which are not full rank, using the triangle inequality in conjunction with the following proposition.

Proposition 2.2 (c.f. Proposition 2.9 [8])**.** *Let X and Y be $\mathbb{R}^d$-valued normal variables with distributions $X \sim \mathrm{MVN}(0, \Sigma)$ and $Y \sim \mathrm{MVN}(0, \Sigma_0)$, where $\Sigma = (\sigma_{i,j})_{i,j=1,\dots,d}$ has full rank, and $\Sigma_0 = (\sigma^0_{i,j})_{i,j=1,\dots,d}$ is non-negative definite. Let $h : \mathbb{R}^d \to \mathbb{R}$ have 3 bounded derivatives. Then*

$$\big|\mathbb{E} h(X) - \mathbb{E} h(Y)\big| \leqslant \frac{1}{2} |h|_2 \sum_{i,j=1}^d |\sigma_{i,j} - \sigma^0_{i,j}|.$$

For non-smooth test functions, following Rinnot and Rotar [9], let Φ denote the standard normal distribution in $\mathbb{R}^d$, and ϕ the corresponding density function. For $h : \mathbb{R}^d \to R$ set

$$h_\delta^+(x) = \sup\{h(x+y) : |y| \leqslant \delta\}, \quad h_\delta^-(x) = \inf\{h(x+y) : |y| \leqslant \delta\},$$

and $\tilde{h}(x, \delta) = h_\delta^+(x) - h_\delta^-(x)$.

Let $\mathcal{H}$ be a class of measurable functions $\mathbb{R}^d \to R$ which are uniformly bounded by 1. Suppose that for any $h \in \mathcal{H}$, for any $\delta > 0$, $h_\delta^+(x)$ and $h_\delta^-(x)$ are in $\mathcal{H}$; for any $d \times d$ matrix A and any vector $b \in \mathbb{R}^d$, $h(Ax + b) \in \mathcal{H}$; and for some constant $a = a(\mathcal{H}, \delta)$, $\sup_{h \in \mathcal{H}} \left\{ \int_{\mathbb{R}^d} \tilde{h}(x, \delta) \Phi(dx) \right\} \leqslant a\delta$. Obviously we may assume $a \geq 1$. The class of indicators of measurable convex sets is such a class where $a \leqslant 2\sqrt{d}$; see the paper by Bolthausen and Götze [1].

Let W have mean vector 0 and covariance matrix Σ. If Λ and R are such that (1.1) is satisfied for W, then $Y = \Sigma^{-1/2} W$ satisfies (1.1) with $\hat{\Lambda} = \Sigma^{-1/2} \Lambda \Sigma^{1/2}$ and $R' = \Sigma^{-1/2} R$. With

$$\hat{\lambda}^{(i)} = \sum_{m=1}^d |(\Sigma^{-1/2} \Lambda^{-1} \Sigma^{1/2})_{m,i}|$$

as well as

(2.4)
$$A' = \sum_{i,j} \hat\lambda^{(i)} \sqrt{\operatorname{Var} \mathbb{E}^Y \sum_{k,\ell} \Sigma_{i,k}^{-1/2} \Sigma_{j,\ell}^{-1/2} (W_k' - W_k)(W_\ell' - W_\ell)},$$

(2.5)
$$B' = \sum_{i,j,k} \hat\lambda^{(i)} \mathbb{E}\biggl| \sum_{r,s,t} \Sigma_{i,r}^{-1/2} \Sigma_{j,s}^{-1/2} \Sigma_{k,t}^{-1/2} (W_r' - W_r)(W_s' - W_s)(W_t' - W_t) \biggr|$$

and

$$\updownarrow abel9 C' = \sum_i \hat\lambda^{(i)} \sqrt{\mathbb{E}\Bigl(\sum_k \Sigma_{i,k}^{-1/2} R_k\Bigr)^2},$$

we have the following result [8].

Corollary 2.3. *Let W be as in Theorem 2.1. Then, for all $h \in \mathcal{H}$ with $|h| \leqslant 1$, there exist $\gamma = \gamma(d)$ and $a > 1$ such that*

$$\sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(Z)| \leqslant \gamma^2 \Bigl(-D' \log(T') + \frac{B'}{2\sqrt{T'}} + C' + a\sqrt{T'} \Bigr),$$

with

$$T' = \frac{1}{a^2} \biggl(D' + \sqrt{\frac{aB'}{2} + D'^2} \biggr)^2 \qquad \textit{and} \qquad D' = \frac{A'}{2} + C'd.$$

Remark 2.4. We can simplify the above bound further. Using Minkowski's inequality we have that $\operatorname{Var} \sum_{i=1}^k X_i \leqslant k^2 \sup_i \operatorname{Var} X_i$, and thus obtain the simple estimate

$$\begin{aligned} &\operatorname{Var} \mathbb{E}^Y \sum_{k,\ell} \Sigma_{i,k}^{-1/2} \Sigma_{j,\ell}^{-1/2} (W_k' - W_k)(W_\ell' - W_\ell) \\ &\qquad \leqslant d^4 \|\Sigma^{-1/2}\|^4 \sup_{k,\ell} \operatorname{Var} \mathbb{E}^W \bigl\{ (W_k' - W_k)(W_\ell' - W_\ell) \bigr\} \end{aligned}$$

and hence

$$A' \leqslant d^3 \|\Sigma^{-1/2}\|^2 \sum_i \hat\lambda^{(i)} \sup_{k,\ell} \sqrt{\operatorname{Var} \mathbb{E}^W \bigl\{ (W_k' - W_k)(W_\ell' - W_\ell) \bigr\}};$$

in B' and C' we could similarly bound $\Sigma_{i,k}^{-1/2}$ by $\|\Sigma^{-1/2}\|$ to obtain a simpler bound. There are however examples, such as the random graph example in Section 5, where $\|\Sigma^{-1/2}\|$ provides a non-informative bound.

Remark 2.5. Note that, if (W, W') is exchangeable and (1.1) is satisfied we have

$$\mathbb{E}(W' - W)(W' - W)^t = 2\mathbb{E}W(\Lambda W)^t = 2\Sigma\Lambda^t \tag{2.6}$$

On the other hand, if we only have $\mathscr{L}(W) = \mathscr{L}(W')$, we obtain

$$\mathbb{E}(W' - W)(W' - W)^t = \Lambda\Sigma + \Sigma\Lambda^t. \tag{2.7}$$

Hence, to check in an application whether the often tedious calculation of Σ and Λ has been carried out correctly, we can combine Equations (2.6) and (2.7), to conclude that, under the conditions of Theorem 2.1, we must have $\Lambda\Sigma = \Sigma\Lambda^t$.

3. The embedding method

Assume that an ℓ-dimensional random vector $W_{(\ell)}$ of interest is given. Often, the construction of an exchangeable pair $(W_{(\ell)}, W'_{(\ell)})$ is straightforward. If, say, $W_{(\ell)} = W_{(\ell)}(\mathbb{X})$ is a function of i.i.d. random variables $\mathbb{X} = (X_1, \dots, X_n)$, one can choose uniformly an index I from 1 to n, replace X_I by an independent copy X'_I, and define $W'_{(\ell)} := W_{(\ell)}(\mathbb{X}')$, where $\mathbb{X}'$ is now the vector $\mathbb{X}$ but with X_I replaced by X'_I.

In general there is no hope that $(W_{(\ell)}, W'_{(\ell)})$ will satisfy Condition (1.1) with R being of the required smaller order or even equal to zero, so that in this case Theorem 2.1 would not yield useful bounds.

Surprisingly often it is possible, though, to extend $W_{(\ell)}$ to a vector $W \in \mathbb{R}^d$ such that we can construct an exchangeable pair (W, W') which satisfies Condition (1.1) with $R = 0$. If we can bound the distance of the distribution $\mathcal{L}(W)$ to a d-dimensional multivariate normal distribution, then a bound on the distance of the distribution $\mathcal{L}(W_{(\ell)})$ to an $\ell-$dimentional multivariate normal distribution follows immediately.

In order to obtain useful bounds in Theorem 2.1, the embedding dimension d should not be too large. In the examples below it will be obvious how to choose $W^{(d-\ell)}$ to make the construction work.

As a first illustration of the method, it was observed in [6] that for functions which depend on the first d coordinates of an infinite Rademacher sequence, that is, a sequence of symmetric $\{-1, 1\}$ random variables, the natural embedding vector is a vector of Rademacher integrals of lower order. A similar construction works fairly generally, as follows. Assume that $F = F(X_1, ..., X_d)$ is a random variable that depends uniquely on the first d coordinates of a sequence X of i.i.d. mean zero random variables, with $E(F) = 0$ and $E(F^2) = 1$, of the form

$$(3.1)\quad F = \sum_{n=1}^{d} \sum_{1 \leqslant i_1 < \dots < i_n \leqslant d} n! f_n(i_1, ..., i_n) X_{i_1} \cdots X_{i_n} =: \sum_{n=1}^{d} J_n(f_n);$$

such representations occur as chaotic decompositions for functionals of Rademacher sequences. A natural exchangeable pair construction is as follows. Pick an index I so that $P(I = i) = \frac{1}{d}$ for $i = 1, \dots, d$, independently of $X_1, ..., X_d$, and if $I = i$ replace X_i by an independent copy X_i^* in all sums in the decomposition (3.1) which involve X_i. Call the resulting expression F', and the corresponding sums $J'_n(f_n); n = 1, \dots, d$. Now choosing as embedding vector $W = (J_1(f_1), \dots, J_d(f_d))$, we check that for all $n = 1, \dots, d$,

$$\begin{aligned} & E(J'_n(f_n) - J_n(f_n)|W) \\ = & -\frac{1}{d} \sum_{i=1}^{d} \sum_{1 \leqslant i_1 < \dots < i_n \leqslant d} \mathbf{1}_{\{i_1, \dots, i_n\}}(i)\ n! f_n(i_1, ..., i_n) E(X_{i_1} \cdots X_{i_n}|W) \\ = & -\frac{n}{d} J_n(f_n). \end{aligned}$$

Thus, with $W' = (J'_1(f_1), \dots, J'_d(f_d))$, the condition (1.1) is satisfied, with $\Lambda = (\lambda_{i,j})_{1 \leqslant i,j \leqslant d}$ being zero off the diagonal and $\lambda_{n,n} = \frac{n}{d}$ for $n = 1, \dots, d$. Note that, although diagonal, the diagonal entries of this Λ are not equal.

It is not possible to correct this by simple coordinate-wise scaling of W as this will change Σ only and leave Λ unaffected; see also the discussion in [8, Section 5]. Hence, again, the generality of (1.1) is essential here.

4. Complete non-degenerate U-statistics

Using the exchangeable pairs coupling, Rinott and Rotar [10] proved a univariate normal approximation theorem for non-degenerate weighted U-statistics with symmetric weight function under fairly mild conditions on the weights. Using the typical coupling, where uniformly a random variable X_i is choosen and replaced by an independent copy, they show that (1.1) is satisfied for the one-dimensional case and a non-trivial remainder term, being Hoeffding projections of smaller order. It should not be difficult (but nevertheless cumbersome) to generalise their result to the multivariate case, where d different U-statistics are regarded based on the same sample of independent random variables, such that (1.1) is satisfied with $\Lambda = I$ and non-trivial remainder term, again of lower order; for multivariate approximations of several U-statistics see also the book by Lee [4]. However, as we want to emphasize the use of Theorem 2.1 for non-diagonal Λ, we take a different approach.

Let $X_1, \dots, X_n$ be a sequence of i.i.d. random elements taking values in a measure space $\mathcal{X}$. Let ψ be a measurable and symmetric function from $\mathcal{X}^d$ to $\mathbb{R}$, and, for each $k = 1, \dots, d$, let

$$\psi_k(X_1, \dots, X_k) := \mathbb{E}\big(\psi(X_1, \dots, X_d) \bigm| X_1, \dots, X_k\big).$$

Assume without loss of generality that $\mathbb{E}\psi(X_1, \dots, X_d) = 0$. For any subset $\alpha \subset \{1, \dots, n\}$ of size k write $\psi_k(\alpha) := \psi(X_{i_1}, \dots, X_{i_k})$ where the i_j are the elements of α. Define the statistics

$$U_k := \sum_{|\alpha|=k} \psi_k(\alpha),$$

where $\sum_{E(\alpha)}$ denotes summation over all subsets $\alpha \subset \{1, \dots, n\}$ which satisfy the property E. Then U_d coincides with the usual U-statistics with kernel ψ. Assume that U_d is non-degenerate, that is, $\mathbb{P}[\psi_1(X_1) = 0] < 1$. Put

$$W_k := n^{1/2} \binom{n}{k}^{-1} U_k.$$

It is well known that $\operatorname{Var} W_k \asymp 1$ (e.g. [4]). Note also that, as $n \to \infty$, $\Sigma := \mathbb{E}(WW^t)$ will converge to a covariance matrix with all entries equal to $\operatorname{Var} \psi_1(X_1)$ and which is thus of rank 1, as we assume non-degeneracy and hence $U_1 = \sum_{i=1}^n \psi_1(X_i)$ will dominate the behaviour of each U_k.

Using Stein's method and the approach of decomposable random variables, Raič [7] proved rates of convergence for vectors of U-statistics, where the coordinates are assumed to be uncorrelated (but nevertheless based upon the same sample $X_1, \dots, X_n$). The next theorem can be seen as a complement to Raič's results, as in our case from above, a normalization is not appropriate.

Theorem 4.1. *With the above notation, and if* $\rho := \mathbb{E}\psi(X_1,\dots,X_d)^4 < \infty$ *we have for every three times differentiable function* h

$$|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2}Z)| \leqslant n^{-\frac{1}{2}}\left(4\rho^{1/2}d^6|h|_2 + \rho^{3/4}d^7|h|_3\right).$$

Proof. Let $X_1',\dots,X_n'$ be independent copies of $X_1,\dots,X_n$. Define the random variables $\psi_{j,k}'(\alpha)$ analogously to $\psi_k(\alpha)$ but based on the sequence $X_1,\dots,X_{j-1},X_j',X_{j+1},\dots,X_n$. Define the coupling as in [10], that is, pick uniformly an index J from $\{1,\dots,n\}$ and replace X_J by X_J', so that $U_k' = \sum_{|\alpha|=k}\psi_{J,k}'(\alpha)$; it is easy to see that (U',U) is exchangeable. Note now that, if $j \notin \alpha$, $\psi_{j,k}'(\alpha) = \psi_k(\alpha)$, and, with $X = (X_1,\dots,X_n)$, that $\mathbb{E}^X\psi_{j,k}'(\alpha) = \psi_{k-1}(\alpha\setminus\{j\})$ if $j \in \alpha$. Thus

$$\begin{aligned}\mathbb{E}^X(U_k'-U_k) &= \frac{1}{n}\sum_{j=1}^n\sum_{\substack{|\alpha|=k,\\ \alpha\ni j}}\mathbb{E}^X\{\psi_{j,k}'(\alpha)-\psi_k(\alpha)\}\\ &= -\frac{k}{n}U_k + \frac{1}{n}\sum_{j=1}^n\sum_{\substack{|\alpha|=k,\\ \alpha\ni j}}\psi_{k-1}(\alpha\setminus\{j\})\\ &= -\frac{k}{n}U_k + \frac{n-k+1}{n}\sum_{|\beta|=k-1}\psi_{k-1}(\beta)\\ &= -\frac{k}{n}U_k + \frac{n-k+1}{n}U_{k-1},\end{aligned} \tag{4.1}$$

where the second last equality follows from the observation that

$$\sum_{\substack{|\alpha|=k,\\ \alpha\ni j}}\psi_{k-1}(\alpha\setminus\{j\}) = \sum_{\substack{|\beta|=k-1,\\ \beta\not\ni j}}\psi_{k-1}(\beta),$$

and thus, in the corresponding double sum of (4.1), every set β of size $k-1$ appears exactly $n-(k-1)$ times. Thus

$$\mathbb{E}^X(W_k'-W_k) = -\frac{k}{n}(W_k - W_{k-1}).$$

Hence, (1.1) is satisfied for $R=0$ and

$$\Lambda = \frac{1}{n}\begin{bmatrix}1 & & & & \\ -2 & 2 & & & \\ & -3 & 3 & & \\ & & \ddots & \ddots & \\ & & & -d & d\end{bmatrix},$$

with lower triangular Λ^{-1} such that, if $l \leqslant k$,

$$(\Lambda^{-1})_{k,l} = n/l,$$

thus, for $l = 1,\dots,d$,

$$\lambda^{(l)} \leqslant dn. \tag{4.2}$$

Define now $\eta_{j,k}(\alpha) := \psi'_{j,k}(\alpha) - \psi_k(\alpha)$. Then we have for every $k,l = 1,\dots,d$,

$$\mathbb{E}^{X,X'}\big\{(U'_k - U_k)(U'_l - U_l)\big\} = \frac{1}{n}\sum_{j=1}^{n}\Bigg(\sum_{\substack{|\alpha|=k,|\beta|=l,\\ \alpha\cap\beta\ni j}} \eta_{j,k}(\alpha)\eta_{j,l}(\beta)\Bigg) \tag{4.3}$$

and

$$\begin{aligned}&\mathbb{E}\big(\mathbb{E}^{X,X'}\big\{(U'_k - U_k)(U'_l - U_l)\big\}\big)^2\\ &\qquad = \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{\substack{|\alpha|=k,|\beta|=l,\\ \alpha\cap\beta\ni i}}\sum_{\substack{|\gamma|=k,|\delta|=l,\\ \gamma\cap\delta\ni j}} \mathbb{E}\big\{\eta_{i,k}(\alpha)\eta_{i,l}(\beta)\eta_{j,k}(\gamma)\eta_{j,l}(\delta)\big\}.\end{aligned} \tag{4.4}$$

Note now that, if the sets $\alpha\cup\beta$ and $\gamma\cup\delta$ are disjoint (which can only happen if $i \neq j$),

$$\mathbb{E}\big\{\eta_{i,k}(\alpha)\eta_{i,k}(\beta)\eta_{j,l}(\gamma)\eta_{j,l}(\delta)\big\} = \mathbb{E}\big\{\eta_{i,k}(\alpha)\eta_{i,k}(\beta)\big\}\mathbb{E}\big\{\eta_{j,l}(\gamma)\eta_{j,l}(\delta)\big\} \tag{4.5}$$

due to independence. The variance of (4.3), that is (4.4) minus the square of the expectation of (4.3), contains only summands where $\alpha\cup\beta$ and $\gamma\cup\delta$ are not disjoint. Recall now that $\rho = \mathbb{E}\psi(X_1,\dots,X_d)^4$. Bounding all the non-vanishing terms simply by 32ρ, it only remains to count the number of non-vanishing terms. Thus,

$$\begin{aligned}&\operatorname{Var}\mathbb{E}^{X,X'}(U'_k - U_k)(U'_l - U_l)\\ &\quad\leqslant \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{\substack{|\alpha|=k,|\beta|=l,\\ \alpha\cap\beta\ni i}}\sum_{\substack{|\gamma|=k,|\delta|=l,\\ \gamma\cap\delta\ni j,(\gamma\cup\delta)\cap(\alpha\cup\beta)\neq\emptyset}} 32\rho\\ &\quad= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{|\alpha|=k,|\beta|=l,\\ \alpha\cap\beta\ni i}}\Bigg(\sum_{j\in\alpha\cup\beta}\sum_{\substack{|\gamma|=k,|\delta|=l,\\ \gamma\cap\delta\ni j}} 32\rho + \sum_{j\notin\alpha\cup\beta}\sum_{\substack{|\gamma|=k,|\delta|=l,\\ \gamma\cap\delta\ni j,(\gamma\cup\delta)\cap(\alpha\cup\beta)\neq\emptyset}} 32\rho\Bigg)\\ &\quad=: A_{k,l} + B_{k,l},\end{aligned}$$

where the equality is just a split of the sum over j into the cases whether or not $j\in\alpha\cup\beta$. In the former case we automatically have $(\alpha\cup\beta)\cap(\gamma\cup\delta)\neq\emptyset$. It is now not difficult to see that

$$A_{k,l} \leqslant \frac{32\rho(k+l-1)}{n}\binom{n-1}{k-1}^2\binom{n-1}{l-1}^2.$$

Noting that, for fixed j, k, l, α and β,

$$\begin{aligned}&\big\{|\gamma| = k, |\delta| = l \,:\, \gamma\cap\delta\ni j, (\gamma\cup\delta)\cap(\alpha\cup\beta)\neq\emptyset\big\}\\ &\qquad = \big\{|\gamma| = k, |\delta| = l \,:\, \gamma\cap\delta\ni j\big\}\\ &\qquad\quad \setminus \big\{|\gamma| = k, |\delta| = l \,:\, \gamma\cap\delta\ni j, (\gamma\cup\delta)\cap(\alpha\cup\beta)=\emptyset\big\},\end{aligned}$$

we further have

$$\begin{aligned}B_{k,l} \leqslant{}& \frac{32\rho(n-1)}{n}\binom{n-1}{k-1}\binom{n-1}{l-1}\times\\ &\times\bigg\{\binom{n-1}{k-1}\binom{n-1}{l-1} - \binom{n-k-l+1}{k-1}\binom{n-k-l+1}{l-1}\bigg\},\end{aligned}$$

where we also used that $\binom{n-|\alpha\cup\beta|}{k-1} \geqslant \binom{n-k-l+1}{k-1}$. The following statements are straightforward to prove:

$$\binom{n-1}{k-1}\binom{n}{k}^{-1} = \frac{k}{n}, \tag{4.6}$$

$$\binom{n-k-l+1}{k-1}\binom{n}{k}^{-1} \geqslant \frac{k}{n}\left(\frac{n-2k-l+3}{n}\right)^{k} \geqslant \frac{k}{n}\left(1-\frac{k(2k+l-3)}{n}\right). \tag{4.7}$$

Thus, from (4.6),

$$n^2\binom{n}{k}^{-2}\binom{n}{l}^{-2}A_{k,l} \leqslant \frac{32\rho(k+l-1)k^2l^2}{n^3} \leqslant \frac{64\rho d^5}{n^3}. \tag{4.8}$$

From (4.6) and (4.7),

$$n^2\binom{n}{k}^{-2}\binom{n}{l}^{-2}B_{k,l} \leqslant \frac{32\rho k^2l^2\big(k(2k+l-3)+l(k+2l-3)\big)}{n^3} \leqslant \frac{192\rho d^6}{n^3}.$$

Thus, for all k and l,

$$\begin{aligned}\operatorname{Var}\mathbb{E}^{W}(W_k'-W_k)(W_l'-W_l) &\leqslant \operatorname{Var}\mathbb{E}^{X,X'}(W_k'-W_k)(W_l'-W_l)\\ &\leqslant \frac{256\rho d^6}{n^3}.\end{aligned} \tag{4.9}$$

Notice further that for any $m=1,\dots,d$,

$$\begin{aligned}\mathbb{E}|U_m'-U_m|^3 &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\Big|\sum_{\substack{|\alpha|=|\beta|=|\gamma|=m\\ \alpha\cap\beta\cap\gamma\ni j}}\eta_{j,m}(\alpha)\eta_{j,m}(\beta)\eta_{j,m}(\gamma)\Big|\\ &\leqslant 8\rho^{3/4}\binom{n-1}{m-1}^{3},\end{aligned}$$

using (5.3); hence, along with (4.6),

$$\begin{aligned}\mathbb{E}|(W_i'-W_i)(W_k'-W_k)(W_l'-W_l)| &\leqslant \max_{m=i,k,l}\mathbb{E}|W_m'-W_m|^3\\ &\leqslant 8\rho^{3/4}n^{3/2}\max_{m=i,k,l}\binom{n}{m}^{-3}\binom{n-1}{m-1}^{3}\\ &\leqslant 8\rho^{3/4}d^3n^{-3/2}.\end{aligned} \tag{4.10}$$

Applying Theorem 2.1 with the estimates (4.2), (4.9) and (4.10) proves the claim. □

Remark 4.2. Using the operator norm as used by Meckes [5] we would be able to achieve a bound of $n\log(d+1)$ instead of (4.2), but using bounds for the total derivatives of the test functions, $\sup_{x\in\mathbb{R}^k}\|D^rh(x)\|_{op}$, instead of bounds for $|h|_r$.

5. Edge and triangle counts in Bernoulli random graphs

Typical summaries for random graphs are the degree distribution and the number of triangles, as a proxy for the clustering coefficient in a random graph, which is the expected ratio of the number of triangles over the number of 2-stars a randomly chosen vertex is involved with. Conditional uniform graph tests are based on fixing the degree distribution and randomising over the edges, conditional on keeping the degree distribution fixed. Our next example shows that even when fixing only the number of edges, not even the degree distribution, under a normal asymptotic regime the number of triangles, and the number of 2-stars, is already asymptotically determined. Let $G(n,p)$ denote a Bernoulli random graph on n vertices, with edge probabilities p; we assume that $n \geq 4$ and that $0 < p < 1$. Let $I_{i,j} = I_{j,i}$ be the Bernoulli(p)-indicator that edge (i,j) is present in the graph; these indicators are independent. Our interest is in the joint distribution of the total number of edges, described by

$$T = \frac{1}{2}\sum_{i,j} I_{i,j} = \sum_{i<j} I_{i,j}$$

and the number of triangles,

$$U = \frac{1}{6}\sum_{i,j,k \text{ distinct}} I_{i,j}I_{j,k}I_{j,k} = \sum_{i<j<k} I_{i,j}I_{j,k}I_{j,k}.$$

Here and in what follows, "i,j,k distinct" is short for "$(i,j,k) : i \neq j \neq k \neq i$"; later we shall also use "i,j,k,ℓ distinct", which is the analogous abbreviation for four indices.

In view of the embedding method, we also include the auxiliary statistic related to the number of 2-stars,

$$V := \frac{1}{2}\sum_{i,j,k \text{ distinct}} I_{i,j}I_{j,k} \;= \sum_{i<j<k} (I_{i,j}I_{j,k} + I_{i,j}I_{i,k} + I_{j,k}I_{i,k}).$$

We note that

$$\mathbb{E}T = \binom{n}{2}p; \qquad \mathbb{E}V = 3\binom{n}{3}p^2, \text{ and} \qquad \mathbb{E}U = \binom{n}{3}p^3.$$

With some calculation, as detailed in Section 5.1, we find that the variances are not all of the same order. Hence, we re-scale our variables (c.f. [3]), putting

$$T_1 = \frac{n-2}{n^2}T, \qquad V_1 = \frac{1}{n^2}V, \qquad \text{and } U_1 = \frac{1}{n^2}U.$$

For these re-scaled variables the covariance matrix Σ_1 for $W_1 = (T_1 - \mathbb{E}T_1, V_1 - \mathbb{E}V_1, U_1 - \mathbb{E}U_1)$ equals
(5.1)

$$\Sigma_1 = 3\frac{(n-2)\binom{n}{3}}{n^4}p(1-p) \times \begin{pmatrix} 1 & 2p & p^2 \\ 2p & 4p^2 + \frac{p(1-p)}{n-2} & 2p^3 + \frac{p^2(1-p)}{n-2} \\ p^2 & 2p^3 + \frac{p^2(1-p)}{n-2} & p^4 + \frac{p^2(1+p-2p^2)}{3(n-2)} \end{pmatrix}.$$

Remark 5.1. With $n \to \infty$ we obtain as approximating covariance matrix

$$\Sigma_0 = \frac{1}{2}p(1-p) \times \begin{pmatrix} 1 & 2p & p^2 \\ 2p & 4p^2 & 2p^3 \\ p^2 & 2p^3 & p^4 \end{pmatrix}. \tag{5.2}$$

As also observed in [3], this matrix has rank 1. It is not difficult to see that the maximal diagonal entry of the inverse Σ^{-1} tends to ∞ as $n \to \infty$, so that a uniform bound on the square root of Σ_1^{-1}, as suggested in Remark 2.4, will not be useful.

Janson and Nowicki [3] derived a normal limit for W_1, but no bounds on the approximation are given. Using Theorem 2.1 we obtain explicit bounds, as follows.

Proposition 5.2. *Let $W_1 = (T_1 - \mathbb{E}T_1, V_1 - \mathbb{E}V_1, U_1 - \mathbb{E}U_1)$ be the centralized count vector of the number of edges, two-stars and triangles in a Bernoulli(p)-random graph. Let Σ_1 be given as in* (5.1)*. Then, for every three times differentiable function h,*

$$\left|\mathbb{E}h(W) - \mathbb{E}h(\Sigma_1^{1/2}Z)\right| \leqslant \frac{|h|_2}{n}\left(\frac{35}{4} + 9n^{-1}\right) + \frac{8|h|_3}{3n}\left(1 + n^{-1} + n^{-2}\right).$$

While we do not claim that the constants in the bound are sharp, as we have $\binom{n}{2}$ random edges in the model, the order $O(n^{-1})$ of the bound is as expected. While for simplicity our other bounds are given as expressions which are uniform in p, bounds dependent on p are derived on the way. In this example, we were not able to obtain any improvement on the bounds using the operator bounds [5].

Proof. The proof consists of two stages. Firstly we construct an exchangeable pair; it will turn out that $R = 0$ in (1.1) and hence C in Theorem 2.1 will vanish. In the second stage we bound the terms A and B in Theorem 2.1.

Construction of an exchangeable pair

Our vector of interest is now $W = (T - \mathbb{E}T, V - \mathbb{E}V, U - \mathbb{E}U)$, re-scaled to $W_1 = (T_1 - \mathbb{E}T_1, V_1 - \mathbb{E}V_1, U_1 - \mathbb{E}U_1)$. We build an exchangeable pair by choosing a potential edge (i,j) uniformly at random, and replacing $I_{i,j}$ by an independent copy $I'_{i,j}$. More formally, pick (I,J) according to

$$\mathbb{P}[I = i, J = j] = \frac{1}{\binom{n}{2}}, \quad 1 \leqslant i < j \leqslant n.$$

If $I = i$, $J = j$ we replace $I_{i,j} = I_{j,i}$ by an independent copy $I'_{i,j} = I'_{j,i}$ and put

$$\begin{aligned}
T' &= T - (I_{I,J} - I'_{I,J}), \\
V' &= V - \sum_{k: k \neq I,J} (I_{I,J} - I'_{I,J})(I_{J,k} + I_{I,k}), \\
U' &= U - \sum_{k: k \neq I,J} (I_{I,J} - I'_{I,J}) I_{J,k} I_{I,k}.
\end{aligned}$$

Put $W' = (T' - \mathbb{E}T, V' - \mathbb{E}V, U' - \mathbb{E}U)$. Then (W, W') forms an exchangeable pair. We re-scale W' as for W to obtain T_1', V_1' and U_1', so that (W_1, W_1') is also exchangeable.

Calculation of Λ

For the conditional expectations $\mathbb{E}^W(W' - W)$, firstly we have

$$\begin{aligned}\mathbb{E}^W(T_1' - T_1) &= \frac{2(n-2)}{n^3(n-1)} \sum_{i<j} \mathbb{E}^W(I_{i,j}' - I_{i,j} | I = i, J = j)\\ &= \frac{n-2}{n^2} p - \frac{2(n-2)}{n^3(n-1)} T = -\frac{1}{\binom{n}{2}} (T_1 - \mathbb{E}T_1).\end{aligned}$$

Furthermore

$$\begin{aligned}-\mathbb{E}^W(V_1' - V_1) \quad &= \frac{1}{n^2\binom{n}{2}} \sum_{i<j} \mathbb{E}^W \sum_{k:k\neq i,j} (I_{i,j} - I_{i,j}')(I_{j,k} + I_{i,k})\\ &= 2\frac{1}{n^2\binom{n}{2}} V - 2p\frac{1}{n^2\binom{n}{2}}(n-2)T\\ &= -2\frac{1}{\binom{n}{2}}(V_1 - \mathbb{E}V_1) + 2p\frac{1}{\binom{n}{2}}(T_1 - \mathbb{E}T_1),\end{aligned}$$

where the last equality follows from $\mathbb{E}(V_1' - V_1) = 0$. Similarly,

$$-\mathbb{E}^W(U_1' - U_1) = -3\frac{1}{\binom{n}{2}}(U_1 - \mathbb{E}U_1) + p\frac{1}{\binom{n}{2}}(V_1 - \mathbb{E}V_1),$$

Using our re-scaling, (1.1) is satisfied with $R = 0$ and Λ given by

$$\Lambda = \frac{1}{\binom{n}{2}} \begin{pmatrix} 1 & 0 & 0 \\ -2p & 2 & 0 \\ 0 & -p & 3 \end{pmatrix}.$$

Bounding A

The inverse matrix Λ^{-1} is easy to calculate; for $\lambda^{(i)} = \sum_{m=1}^d |(\Lambda^{-1})_{m,i}|$, for simplicity we shall apply the uniform bound

$$|\lambda^{(i)}| \leqslant \frac{3}{2} n^2, \quad i = 1, 2, 3.$$

As the bounding of the conditional variances is somewhat laborious, most of the work can be found in Section 5.2. The conditional variances involving $T' - T$ can be calculated exactly. As $I_{i,j}^2 = I_{i,j}$,

$$\begin{aligned}\mathbb{E}^W(T' - T)^2 &= \frac{1}{\binom{n}{2}} \sum_{i<j} \mathbb{E}^W(I_{i,j}' - I_{i,j})^2\\ &= \frac{1}{\binom{n}{2}} \sum_{i<j} \{p - p\mathbb{E}^W I_{i,j} + (1-p)\mathbb{E}^W I_{i,j}\} = p + (1-2p)\frac{1}{\binom{n}{2}} T,\end{aligned}$$

so that with $\operatorname{Var} T$ given in (5.5), $\operatorname{Var}(\mathbb{E}^W(T' - T)^2) = \frac{1}{\binom{n}{2}}(1-2p)^2 p(1-p)$ and

$$\operatorname{Var}(\mathbb{E}^W(T_1' - T_1)^2) = \frac{(n-2)^4}{n^8\binom{n}{2}}(1-2p)^2 p(1-p) < n^{-6},$$

where we used that $p(1-p) \leqslant 1/4$ for all p. Thus

$$\sqrt{\operatorname{Var}(\mathbb{E}^W(T_1'-T_1)^2))} < n^{-3}.$$

Similarly we obtain

$$\sqrt{\operatorname{Var}\mathbb{E}^W(T_1'-T_1)(V_1'-V_1)} < 2n^{-3}$$

and

$$\sqrt{\operatorname{Var}\mathbb{E}^W(T_1'-T_1)(U_1'-U_1)} < 2n^{-3}.$$

The conditional variances involving V and U only are more involved; the calculations are available in Section 5.2. We obtain after some calculations that $\operatorname{Var}(\mathbb{E}^W(V'-V)^2) \leqslant 33n^2$, and hence

$$\sqrt{\operatorname{Var}(\mathbb{E}^W(V_1'-V_1)^2)} < 6n^{-3}.$$

For $\mathbb{E}^W(V_1'-V_1)(U_1'-U_1)$, analogous calculations lead to

$$\sqrt{\operatorname{Var}\left(\mathbb{E}^W(V_1'-V_1)(U_1'-U_1)\right)} < n^{-3} + 11n^{-4}.$$

Finally, again using our variance inequalities,

$$\sqrt{\operatorname{Var}\left(\mathbb{E}^W(U_1'-U_1)^2\right)} < 5n^{-3} + 2n^{-4}.$$

Collecting these bounds we obtain for A in Theorem 2.1 that

$$A < 35n^{-1} + 36n^{-2}.$$

Bounding B

We use the generalized Hölder inequality

$$\mathbb{E}\prod_{i=1}^{3}|X_i| \leqslant \prod_{i=1}^{3}\{\mathbb{E}|X_i|^3\}^{\frac{1}{3}} \leqslant \max_{i=1,2,3}\mathbb{E}|X_i|^3. \tag{5.3}$$

Again the complete calculations are found in Section 5.3. To illustrate the calculation,

$$\mathbb{E}|T'-T|^3 = \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}|I_{i,j}-I_{i,j}'|^3 = 2p(1-p) \leqslant \frac{1}{2},$$

so that

$$\mathbb{E}|T_1'-T_1|^3 = \frac{(n-2)^3}{n^6}2p(1-p) < \frac{1}{2}n^{-3}.$$

Similar calculations yield ,

$$\mathbb{E}|V_1'-V_1|^3 < \frac{64}{27}\left(n^{-3}+n^{-4}+n^{-5}\right),$$

as well as

$$\mathbb{E}|U_1'-U_1|^3 < \frac{27}{128}\left(n^{-3}+n^{-4}+n^{-5}\right).$$

Thus for B in Theorem 2.1 we have

$$B < \frac{3}{2}n^2 \times 9 \times \frac{64}{27}\left(n^{-3}+n^{-4}+n^{-5}\right) = 32\left(n^{-1}+n^{-2}+n^{-3}\right).$$

Collecting the bounds gives the result. □

Remark 5.3. Had we not introduced V, conditioning would yield

$$\begin{aligned}-\mathbb{E}^{T,U}(U'-U) &= \frac{2}{n(n-1)}\sum_{i<j}\mathbb{E}^{T,U}\sum_{k:k\neq i,j}(I_{i,j}I_{j,k}I_{i,k}-I'_{i,j}I_{j,k}I_{i,k})\\ &= 3\frac{2}{n(n-1)}U - p\frac{2}{n(n-1)}\mathbb{E}^{T,U}\sum_{i<j,\,k\neq i,j}I_{j,k}I_{i,k}.\end{aligned}$$

The expression $\sum_{i<j,\,k\neq i,j}\mathbb{E}^{T,U}I_{j,k}I_{j,k}$ would result in a non-linear remainder term R in Equation (1.1). The introduction of V not only avoids this remainder term, indeed $R=0$ in (1.1), but also yields a more detailed result. This observation that the 2-stars form a useful auxiliary statistic can also be found in [3]; there it is related to Hoeffding-type projections.

Using Proposition 2.2, we also obtain a normal approximation for Σ_0 given in (5.2).

Corollary 5.4. *Under the assumptions of Proposition 5.2, for every three times differentiable function h,*

$$\begin{aligned}\left|\mathbb{E}h(W)-\mathbb{E}h(\Sigma_0^{1/2}Z)\right| &\leqslant \frac{|h|_2}{2n}\left(44+21n^{-1}+32n^{-2}+4n^{-3}\right)\\ &\quad+\frac{8|h|_3}{3n}\left(1+n^{-1}+n^{-2}\right).\end{aligned}$$

Proof. We employ Proposition 5.2 and Proposition 2.2, with the triangle inequality. A straightforward calculation shows that $\left|\frac{3(n-2)\binom{n}{3}}{n^4}-\frac{1}{2}\right|\leqslant\frac{3}{2}n^{-1}+2n^{-3}$ and so

$$\begin{aligned}&\sum_{i,j=1}^{d}|\sigma_{i,j}-\sigma^0_{i,j}|\\ &\leqslant\left(\frac{3}{2}n^{-1}+2n^{-3}\right)\left\{1+4p+6p^2+4p^3+p^4\right\}\\ &\quad+\left(\frac{p(1-p)}{n-2}+2\frac{p^2(1-p)}{n-2}+\frac{p^2(1-p)(4-p)}{3(n-2)}\right)\left(\frac{3}{2}n^{-1}+2n^{-3}+1\right)\\ &< 26n^{-1}+3n^{-2}+32n^{-3}+4n^{-4}.\end{aligned}$$

Here we used the crude bound that $(n-2)^{-1}\leqslant\frac{3}{2}n^{-1}$. The corollary follows. □

5.1. **Calculation of the covariance matrix.** To calculate the covariance matrix Σ, we put

$$\tilde{I}_{i,j}=I_{i,j}-p$$

as the centralized edge indicator, and similarly we centralize

$$\tilde{T}=\sum_{i<j}\tilde{I}_{i,j},\quad \tilde{V}=\frac{1}{2}\sum_{i,j,k\text{ distinct}}\tilde{I}_{i,j}\tilde{I}_{j,k}\text{ and }\tilde{U}=\sum_{i<j<k}\tilde{I}_{i,j}\tilde{I}_{j,k}\tilde{I}_{i,k}.$$

Then, by independence, all these quantities have mean zero. For the variances, the expectation of the product of centralized indicators vanish unless

all the centralized indicators involved are raised to an even power. Hence
(5.4)

$$\operatorname{Var}\tilde{T} = \binom{n}{2}p(1-p), \quad \operatorname{Var}\tilde{V} = 3\binom{n}{3}p^2(1-p)^2, \operatorname{Var}\tilde{U} = \binom{n}{3}p^3(1-p)^3.$$

Moreover, for the same reason, all covariances between the centralised variables vanish. Expressing T, V and U, we have $\tilde{T} = T - \mathbb{E}T$ so that

$$T = \tilde{T} + \mathbb{E}T = \tilde{T} + \binom{n}{2}p$$

and

$$\operatorname{Var}T = \binom{n}{2}p(1-p) = 3\binom{n}{3}\frac{1}{n-2}p(1-p). \tag{5.5}$$

Next, $\tilde{V} = V - 2p(n-2)T + 3p^2\binom{n}{3}$, so that

$$V = \tilde{V} + 2(n-2)p\tilde{T} + 3\binom{n}{3}p^2.$$

As $\tilde{V}$ and $\tilde{T}$ are uncorrelated, this gives that

$$\operatorname{Var}V = 3\binom{n}{3}p^2(1-p)\{1-p+4(n-2)p\}. \tag{5.6}$$

For U, we have $\tilde{U} = U - pV + p^2(n-2)T - p^3\binom{n}{3}$. Using the above expressions (5.1) and (5.1) for T and V we obtain

$$U = \tilde{U} + p\tilde{V} + p^2(n-2)\tilde{T} + p^3\binom{n}{3}.$$

This gives for the variance

$$\operatorname{Var}U = \binom{n}{3}p^3(1-p)\left\{(1-p)^2 + 3p(1-p) + 3(n-2)p^2\right\}. \tag{5.7}$$

We can now also calculate the covariances. Again we use that the centralized variables are uncorrelated to obtain

$$\operatorname{Cov}(T,V) = \operatorname{Cov}\left(\tilde{T}, \tilde{V} + 2(n-2)p\tilde{T}\right) = 2(n-2)p\operatorname{Var}(\tilde{T}) = 6\binom{n}{3}p^2(1-p).$$

Similarly, we calculate that $\operatorname{Cov}(T,U) = 3\binom{n}{3}p^3(1-p)$, and $\operatorname{Cov}(V,U) = 3\binom{n}{3}p^3(1-p)\,(1-p+2(n-2)p)$. Re-scaling gives the covariance matrix (5.1).

5.2. **Calculation of the conditional variances.** For the conditional variances in the random graph example, the calculations are somewhat involved. We repeat the first calculation in more detail before moving on to further bounds. With (5.1),

$$\begin{aligned}
&\mathbb{E}^W(T'-T)(V'-V) \\
&= -\frac{1}{\binom{n}{2}} \sum_{i<j,\, k\neq i,j} \mathbb{E}^W(I'_{i,j} - I_{i,j})^2(I_{i,k} + I_{j,k}) \\
&= \frac{1}{\binom{n}{2}}(-2(n-2)pT - 2(1-2p)V)
\end{aligned}$$

so that

$$\begin{aligned}&\operatorname{Var}\mathbb{E}^W(T'-T)(V'-V)=\\&\qquad\frac{4(n-2)}{\binom{n}{2}}p(1-p)\big\{(n-2)p^2(3-4p)^2+(1-2p)^2p(1-p)\big\}<4,\end{aligned}$$

where we used that $p^3(1-p)\leqslant\frac{27}{256}$ and that $n\geq 4$. Thus

$$\sqrt{\operatorname{Var}\mathbb{E}^W(T'-T)(V'-V)}<2n^{-3}.$$

Similarly, with (5.1),

$$\begin{aligned}&\mathbb{E}^W(T'-T)(U'-U)\\&\quad=\frac{1}{\binom{n}{2}}\sum_{i<j,\,k\neq i,j}\big\{p\mathbb{E}^W I_{j,k}I_{i,k}+(1-2p)\mathbb{E}^W I_{i,j}I_{j,k}I_{i,k}\big\}\\&\quad=\frac{1}{\binom{n}{2}}\big(pV+3(1-2p)U\big).\end{aligned}$$

Thus we calculate that

$$\begin{aligned}&\operatorname{Var}\mathbb{E}^W(T'-T)(U'-U)\\=&\frac{n-2}{\binom{n}{2}}p^3(1-p)\left(3(1-2p)^2(1-p)^2+p(1-p)(4-6p)^2+(n-2)p^2(5-6p)^2\right)\end{aligned}$$

and, using that $p(5-6p)\leqslant\frac{25}{24}$ and $p^3(1-p)\leqslant\frac{27}{256}$, we obtain

$$\sqrt{\operatorname{Var}\mathbb{E}^W(T'-T)(U'-U)}<n^{-3}.$$

For $\operatorname{Var}\mathbb{E}^W(V'-V)^2$ we introduce the notation

$$N_i=\sum_{j:j\neq i}I_{i,j},\qquad M_{i,j}=\sum_{k:k\neq i,j}I_{i,k}I_{k,j}. \tag{5.8}$$

Then

$$T=\frac{1}{2}\sum_i N_i, \tag{5.9}$$

$$V=\frac{1}{2}\sum_{i\neq j}M_{i,j}=\frac{1}{2}\sum_{i\neq j}I_{i,j}N_i-T=\frac{1}{2}\sum_i N_i^2-T, \tag{5.10}$$

$$U=\frac{1}{6}\sum_{i\neq j}I_{i,j}M_{i,j}. \tag{5.11}$$

We have

$$\begin{aligned}
&\mathbb{E}^W(V'-V)^2\\
&\quad= \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}^W(I_{i,j}-I'_{i,j})^2\,(N_j+N_i-2I_{i,j})^2\\
&\quad= \frac{1}{2\binom{n}{2}}\sum_{i\neq j}\Big\{p\mathbb{E}^W\,(N_j+N_i-2I_{i,j})^2+(1-2p)\mathbb{E}^W I_{i,j}\,(N_j+N_i-2I_{i,j})^2\Big\}\\
&\quad= \frac{1}{2\binom{n}{2}}\Bigg\{p\mathbb{E}^W\,\big(4(n-2)(V+T)-8T+8T^2-16V)\big)\\
&\qquad\qquad+(1-2p)\mathbb{E}^W\Bigg(2\sum_{i\neq j}I_{i,j}N_i^2-8T+2\sum_{i\neq j}I_{i,j}N_iN_j-16V)\Bigg)\Bigg\},
\end{aligned}$$

where we used (5.9) and (5.10) for the last equation. To simplify this expression, note that $\sum_i N_i^2 = 2T+2V$, and $\sum_{i\neq j}N_iN_j = 4T^2-2T-2V$ as well as

$$\sum_{i\neq j}I_{i,j}N_i^2 = \sum_{i,j,k,\ell\text{ distinct}} I_{i,j}I_{i,k}I_{i,\ell}+6V+2T,$$

and

$$\sum_{i\neq j}I_{i,j}N_iN_j = \sum_{i,j,k,\ell\text{ distinct}} I_{i,j}I_{i,k}I_{j,\ell}+4V+6U+2T,$$

so that

$$\begin{aligned}
\mathbb{E}^W(V'-V)^2 = \frac{1}{\binom{n}{2}}\Bigg\{&2p(n-4)T+2V(np-10p+2)+6(1-2p)U+4pT^2\\
&+(1-2p)\sum_{\substack{i,j,k,\ell\\ \text{distinct}}}\mathbb{E}^W I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell})\Bigg\}.
\end{aligned}$$

With the notation $\tilde{T}$ for the centralized variable, we have that

$$\begin{aligned}
&\operatorname{Var}\mathbb{E}^W(V'-V)^2\\
&\quad\leqslant 5\frac{1}{\binom{n}{2}^2}\Bigg\{p^2(2n-8+4pn^2-4pn)^2\operatorname{Var}(T)+4(np-10p+2)^2\operatorname{Var}(V)\\
&\qquad\qquad+36(1-2p)^2\operatorname{Var}(U)+16p^2\operatorname{Var}(\tilde{T}^2)\\
&\qquad\qquad+(1-2p)^2\operatorname{Var}\Big(\sum_{\substack{i,j,k,\ell\\ \text{distinct}}}\mathbb{E}^W I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell})\Big)\Bigg\},
\end{aligned}$$

where we used that in general $\operatorname{Var}\sum_{i=1}^k X_i \leqslant k\sum_{i=1}^k \operatorname{Var} X_i$ and (5.4). Here, the variances for T, V and U are given in (5.5), (5.6), and (5.7). To simplify the expression, we use that $p^3(1-p)\leqslant 27/256$ to bound

$$p^2(2n-8+4pn^2-4pn)^2\operatorname{Var}(T)\leqslant \frac{27}{64}\binom{n}{2}n^2(n+2)^2.$$

Similarly, we bound with $p^2(1-p) \leqslant 4/27$ and $n \geq 4$

$$4(np-10p+2)^2 \operatorname{Var}(V) \leqslant \frac{16}{27} n^3(n-1)(n-2)(n+1),$$

and

$$36(1-2p)^2 \operatorname{Var}(U) \leqslant \frac{81}{256} n(n-1)(n-2)(3n+2).$$

We note that $\mathbb{E}\tilde{I}_{i,j}\tilde{I}_{u,v}\tilde{I}_{s,t}\tilde{I}_{k,\ell} = 0$ unless either all pairs of indices are the same, or the product is made up of two distinct index pairs only. Hence

$$\begin{aligned}
\operatorname{Var}\tilde{T}^2 &= \sum_{i<j}\sum_{u<v}\sum_{s<t}\sum_{k<\ell} \mathbb{E}\tilde{I}_{i,j}\tilde{I}_{u,v}\tilde{I}_{s,t}\tilde{I}_{k,\ell} \\
&= \binom{n}{2} p(1-p)\left\{3\left(\binom{n}{2}-1\right)p(1-p)+(1-p)^3+p^3\right\} \\
&< n^2\binom{n}{2}p(1-p),
\end{aligned}$$

giving

$$16p^2 \operatorname{Var}\tilde{T}^2 \leqslant \frac{27}{32} n^3(n-1).$$

For the last variance term, we use that conditional variances can be bounded by unconditional variances, giving

$$\begin{aligned}
&\operatorname{Var}\sum_{i\neq j}\sum_{k:k\neq i,j}\sum_{\ell:\ell\neq i,j,k} \mathbb{E}^W I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell}) \\
&\leqslant \sum_{\substack{i,j,k,\ell\\ \text{distinct}}} \operatorname{Var} I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell}) \\
&\qquad + \sum_{i,j,k,\ell \text{ distinct}}\ \sum_{r,s,t,u \text{ distinct}} \mathbf{1}((i,j,k,\ell)\neq(r,s,t,u)) \\
&\qquad \times \mathbf{1}(|\{i,j,k,\ell\}\cap\{r,s,t,u\}|\geq 2) \\
&\qquad \times \operatorname{Cov}(I_{i,j}I_{i,k}(I_{i,\ell}+I_{j,\ell}), I_{r,s}I_{r,t}(I_{r,u}+I_{s,u}))) \\
&\leqslant 2\binom{n}{4}\left(p^3(1-p^3)+4\binom{4}{2}\binom{n}{2}p^2(1-p^4)\right) < 3n^2\binom{n}{4}.
\end{aligned}$$

Here we used the independence of the edge indicators. For the last bound we employed that $p^3(1-p^3) \leqslant 1/4$, that $p^2(1-p^4) \leqslant (\sqrt{3}-1)/3$, and that $n \geq 4$. Collecting the variances and using that $n \geq 4$,

$$\begin{aligned}
&\operatorname{Var}(\mathbb{E}^W(V'-V)^2) \\
&\leqslant 5\frac{1}{\binom{n}{2}^2}\left\{\frac{27}{64}\binom{n}{2}n^2(n+2)^2+\frac{16}{27}n^3(n-1)(n-2)(n+1)\right. \\
&\qquad \left. +\frac{81}{256}n(n-1)(n-2)(3n+2)+\frac{27}{32}n^3(n-1)+3n^2\binom{n}{4}\right\}
\end{aligned}$$

This gives that

$$\sqrt{\operatorname{Var}(\mathbb{E}^W(V_1'-V_1)^2)} < 6n^{-3}.$$

For $\mathbb{E}^W(V'-V)(U'-U)$, we have,

$$\begin{aligned}&\mathbb{E}^W(V'-V)(U'-U)\\&\quad=\frac{1}{\binom{n}{2}}\Big(p\sum_{i\neq j}\mathbb{E}^W N_iM_{i,j}-6(1-p)U+(1-2p)\sum_{i\neq j}\mathbb{E}^W I_{i,j}N_iM_{i,j}\Big).\end{aligned}$$

where we used (5.11). Now

$$\sum_{i\neq j}N_iM_{i,j}=2V+6U+\sum_{\substack{i,j,k,\ell\\\text{distinct}}}I_{i,k}I_{k,j}I_{i,\ell}.$$

Similarly,

$$\sum_{i\neq j}I_{i,j}N_iM_{i,j}=12U+\sum_{\substack{i,j,k,\ell\\\text{distinct}}}I_{i,j}I_{i,k}I_{i,\ell}I_{\ell,j},$$

so that

$$\begin{aligned}\mathbb{E}^W(V'-V)(U'-U)=\frac{1}{\binom{n}{2}}\Big(&2pV+6(1-2p)U+p\sum_{\substack{i,j,k,\ell\\\text{distinct}}}I_{i,k}I_{k,j}I_{i,\ell}\\&+(1-2p)\sum_{\substack{i,j,k,\ell\\\text{distinct}}}I_{i,j}I_{i,k}I_{i,\ell}I_{\ell,j}\Big).\end{aligned}$$

Furthermore, as before,

$$\begin{aligned}&\operatorname{Var}\sum_{i,j,k,\ell\text{ distinct}}\mathbb{E}^W I_{i,k}I_{k,j}I_{i,\ell}\\&\quad\leqslant\binom{n}{4}\left(p^3(1-p^3)+6\binom{n}{2}p^2(1-p^4)\right)<\binom{n}{4}n^2.\end{aligned}$$

Similarly as for (5.2),

$$\begin{aligned}\operatorname{Var}\sum_{\substack{i,j,k,\ell\\\text{distinct}}}\mathbb{E}^W I_{i,j}I_{i,k}I_{i,\ell}I_{j,\ell}&\leqslant\operatorname{Var}\sum_{\substack{i,j,k,\ell\\\text{distinct}}}I_{i,j}I_{i,k}I_{i,\ell}I_{j,\ell}\\&\leqslant\binom{n}{4}\left(p^4(1-p^4)+6\binom{n}{2}p^2(1-p^6)\right)\\&<\binom{n}{4}\left(\frac{1}{256}+\frac{1}{16}\binom{n}{2}\right).\end{aligned}$$

As $p<1$, we obtain that

$$\begin{aligned}&\operatorname{Var}\mathbb{E}^W(V'-V)(U'-U)\\&\quad<4\frac{1}{\binom{n}{2}^2}\bigg\{12\frac{27}{256}\binom{n}{3}\big(16(n-2)+1)+9n+9\big)\\&\qquad\qquad+\binom{n}{4}n^2+\binom{n}{4}\left(\frac{1}{256}+\frac{1}{16}\binom{n}{2}\right)\bigg\}<n^2+108\end{aligned}$$

so that

$$\sqrt{\operatorname{Var}\left(\mathbb{E}^W(V_1'-V_1)(U_1'-U_1)\right)}<n^{-3}+11n^{-4}.$$

Finally,

$$\mathbb{E}^W(U'-U)^2 = \frac{1}{2\binom{n}{2}}\sum_{i\neq j}\bigl(p\mathbb{E}^W M_{i,j}^2 + (1-2p)\mathbb{E}^W I_{i,j}M_{i,j}^2\bigr).$$

We have that

$$M_{i,j}^2 = \sum_{k:k\neq i,j}\sum_{\ell:\ell\neq i,j} I_{i,k}I_{k,j}I_{i,\ell}I_{\ell,j} = M_{i,j} + \sum_{k:k\neq i,j}\sum_{\ell:\ell\neq i,j,k} I_{i,k}I_{k,j}I_{i,\ell}I_{\ell,j},$$

and

$$I_{i,j}M_{i,j}^2 = I_{i,j}M_{i,j} + \sum_{k:k\neq i,j}\sum_{\ell:\ell\neq i,j,k} I_{i,j}I_{i,k}I_{k,j}I_{i,\ell}I_{\ell,j},$$

so that

$$\begin{aligned}\mathbb{E}^W(U'-U)^2 = \frac{1}{2\binom{n}{2}}\Biggl\{&2pV + 6(1-2p)U + p\sum_{\substack{i,j,k,\ell\\ \text{distinct}}}\mathbb{E}^W I_{i,k}I_{k,j}I_{i,\ell}I_{\ell,j}\\ &+ (1-2p)\sum_{\substack{i,j,k,\ell\\ \text{distinct}}}\mathbb{E}^W I_{i,j}I_{i,k}I_{k,j}I_{i,\ell}I_{\ell,j}\Biggr\}.\end{aligned}$$

As for (5.2), we obtain

$$\operatorname{Var}\sum_{\substack{i,j,k,\ell\\ \text{distinct}}}\mathbb{E}^W I_{i,k}I_{k,j}I_{i,\ell}I_{j,\ell} \leqslant \binom{n}{4}\left(p^4(1-p^4) + 6\binom{n}{2}p^2(1-p^6)\right)$$

and

$$\operatorname{Var}\sum_{\substack{i,j,k,\ell\\ \text{distinct}}}\mathbb{E}^W I_{i,j}I_{i,k}I_{k,j}I_{i,\ell}I_{j,\ell} \leqslant \binom{n}{4}\left(p^5(1-p^5) + 6\binom{n}{2}p^2(1-p^8)\right).$$

Again using our variance inequalities, we thus obtain that

$$\begin{aligned}&\operatorname{Var}\bigl(\mathbb{E}^W(U'-U)^2\bigr)\\ &\quad\leqslant \frac{1}{\binom{n}{2}^2}\Biggl\{3\binom{n}{3}p^3(1-p)\biggl(4p(4(n-2)p+1-p)\\ &\qquad\qquad + 36(1-2p)^2((n-2)p^2 + \frac{1}{3}(4-5p+p^2))\biggr)\\ &\qquad + p^2\binom{n}{4}\left(p^4(1-p^4) + 6\binom{n}{2}p^2(1-p^6)\right)\\ &\qquad + (1-2p)^2\binom{n}{4}\left(p^5(1-p^5) + 6\binom{n}{2}p^2(1-p^8)\right)\Biggr\}\\ &\quad\leqslant 22 + 2n^2,\end{aligned}$$

so that

$$\sqrt{\operatorname{Var}\bigl(\mathbb{E}^W(U'-U)^2\bigr)} < 5n^{-3} + 2n^{-4}.$$

5.3. **Calculation of the third moments.** Firstly, $\mathbb{E}|T'-T|^3 = \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}|I_{i,j}-I'_{i,j}|^3 = 2p(1-p) < \frac{1}{2}$, so that

$$\mathbb{E}|T_1'-T_1|^3 = \frac{(n-2)^3}{n^6}2p(1-p) < \frac{1}{2}n^{-3}.$$

Similarly,

$$\begin{aligned}
&\mathbb{E}|V'-V|^3 \\
&\quad= \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}|I_{i,j}-I'_{i,j}|^3 \sum_{k,\ell,s:k,\ell,s\neq i,j}(I_{j,k}+I_{i,k})(I_{j,\ell}+I_{i,\ell})(I_{j,s}+I_{i,s}) \\
&\quad= 2p(1-p)(n-2)\times \\
&\qquad\times\left(8p^2+2p(1-p)+2(n-3)(2p^2+2p^3)+8(n-3)(n-4)p^3\right),
\end{aligned}$$

so that

$$\mathbb{E}|V_1'-V_1|^3 < \frac{64}{27}\left(n^{-3}+n^{-4}+n^{-5}\right).$$

Lastly,

$$\begin{aligned}
\mathbb{E}|U'-U|^3 &= \frac{1}{\binom{n}{2}}\sum_{i<j}\mathbb{E}|I_{i,j}-I'_{i,j}|^3\sum_{k:k\neq i,j}\sum_{\ell:\ell\neq i,j}\sum_{s:s\neq i,j}I_{j,k}I_{i,k}I_{j,\ell}I_{i,\ell}I_{j,s}I_{i,s} \\
&= 2p(1-p)(n-2)\left(p^2+(n-3)p^4+(n-3)(n-4)p^6\right),
\end{aligned}$$

so that

$$\mathbb{E}|U_1'-U_1|^3 < \frac{54}{256}\left(n^{-3}+n^{-4}+n^{-5}\right).$$

References

[1] Bolthausen, E. and Götze, F. (1993). The rate of convergence for multivariate sampling statistics. *Ann. Statist.* **21,** 1692–1710.

[2] Chatterjee, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138,** 305–321.

[3] Janson, S. and Nowicki, K. (1991). The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probab. Theory Related Fields* **90,** 341–375.

[4] Lee, A. J. (1990). *U-statistics* vol. 110 of *Statistics: Textbooks and Monographs.* Marcel Dekker Inc., New York. Theory and practice.

[5] Meckes, E. S. (2009). On Stein's method for multivariate normal approximation. *to appear in High Dimensional Probability V*.

[6] Nourdin, I., Peccati, G. and Reinert, G. Stein's method and stochastic analysis of Rademacher functionals. arXiv:0810.2890v3 (2009).

[7] Raič, M. (2004). A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Probab.* **17,** 573–603.

[8] Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37,** 2150–2173.

[9] Rinott, Y. and Rotar, V. (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.* **56,** 333–350.

[10] Rinott, Y. and Rotar, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7,** 1080–1105.

[11] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory.* Univ. California Press, Berkeley, Calif. pp. 583–602.

[12] Stein, C. (1986). *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA.